PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-233389
(43)Date of publication of application : 22.08.2003

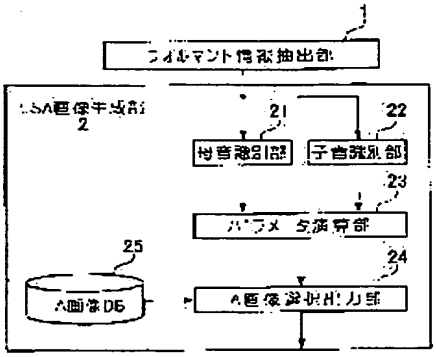
(51)Int.Cl. G10L 15/00
G06T 11/60
G06T 13/00
G10L 15/02
G10L 15/10

(21)Application number : 2002-034465
(22)Date of filing : 12.02.2002

(71)Applicant : YAMAHA CORP
(72)Inventor : NISHIMOTO TETSUO

(54) ANIMATION IMAGE GENERATING DEVICE, PORTABLE TELEPHONE HAVING THE DEVICE INSIDE, AND ANIMATION IMAGE GENERATING METHOD

(57)Abstract:
PROBLEM TO BE SOLVED: To generate a lip sink animation image from a real-time speech or a speech having been sound-recorded without depending upon a program.
SOLUTION: An animation image generating device comprises a formant information extraction part 1 which extracts formant information featuring vowels by analyzing a speech, and a lip sink animation image generation part 2 which generates the lip sink animation image following up variation of the extracted formant information. The lip sink animation image generation part 2 comprises a vowel identification part 21 which identifies the vowels from the relative positional relation of the extracted low-order formant frequencies, and a lip animation image selective output part 24 which selects and outputs a prepared lip animation image in synchronism with the identified vowels.



LEGAL STATUS

[Date of request for examination] 20.11.2003
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－233389
（P2003－233389A）
(43)公開日　平成15年８月22日(2003.8.22)

(51)Int.Cl.'	識別記号	FI		テーマコード'(参考)
Ｇ１０Ｌ　15/00		Ｇ０６Ｔ　11/60	２００	５Ｂ０５０
Ｇ０６Ｔ　11/60	２００	13/00	Ａ	５Ｄ０１５
13/00		Ｇ１０Ｌ　3/00	５５１Ｇ	
Ｇ１０Ｌ　15/02		9/02	３０１Ａ	
15/10		3/00	５３１Ｚ	

審査請求　未請求　請求項の数７　ＯＬ　（全７頁）　最終頁に続く

(21)出願番号　特願2002－34465(P2002－34465)

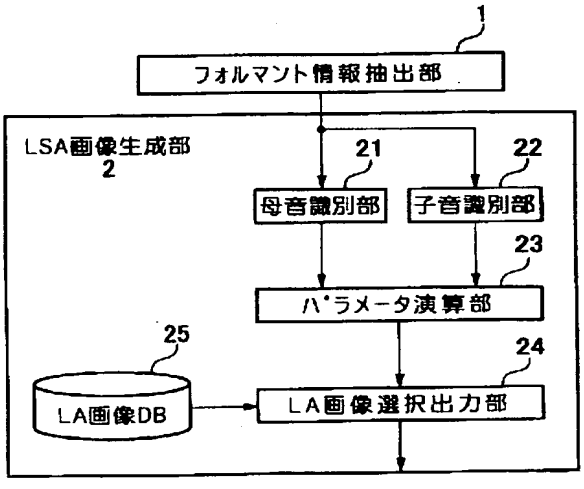
(22)出願日　平成14年２月12日(2002.2.12)

(71)出願人　000004075
ヤマハ株式会社
静岡県浜松市中沢町10番１号
(72)発明者　西元　哲夫
静岡県浜松市中沢町10番１号　ヤマハ株式会社内
(74)代理人　100064908
弁理士　志賀　正武　（外１名）

Ｆターム(参考)　5B050 BA07 BA08 BA12 EA19 EA24
FA02 FA10 FA19
5D015 KK01 KK02

(54)【発明の名称】　アニメーション画像生成装置、及び同装置を内蔵した携帯電話、並びにアニメーション画像生成方法

(57)【要約】
【課題】　プログラムに依存することなく実時間音声あるいは既録音の音声からリップシンク・アニメーション画像を生成する。
【解決手段】　本発明のアニメーション画像生成装置は、音声を分析して母音を特徴づけるフォルマント情報を抽出するフォルマント情報抽出部１と、抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するリップシンク・アニメーション画像生成部２から成る。また、リップシンク・アニメーション画像生成部２は、抽出される低次のフォルマント周波数の相対的位置関係により母音を識別する母音識別部２１と、識別された母音に同期してあらかじめ用意されたリップアニメーション画像を選択出力するリップアニメーション画像選択出力部２４から成る。



【特許請求の範囲】

【請求項１】　音声を分析して母音を特徴づけるフォルマント情報を抽出するフォルマント情報抽出手段と、
前記抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するリップシンク・アニメーション画像生成手段と、を備えたことを特徴とするアニメーション画像生成装置。

【請求項２】　前記リップシンク・アニメーション画像生成手段は、
前記フォルマント情報抽出手段により抽出される低次のフォルマント周波数の相対的位置関係により母音を識別する母音識別手段と、
前記識別された母音に同期してあらかじめ用意されたリップアニメーション画像を選択出力するリップアニメーション画像選択出力手段と、を備えたことを特徴とする請求項１に記載のアニメーション画像生成装置。

【請求項３】　前記リップシンク・アニメーション画像生成手段は、
子音を識別し、前記識別された子音に同期してあらかじめ用意されたリップアニメーション画像を選択出力するリップアニメーション画像選択出力手段、を備えたことを特徴とする請求項１または２に記載のアニメーション画像生成装置。

【請求項４】　着信音声を分析して母音を特徴づけるフォルマント情報を抽出するフォルマント情報抽出手段と、
着信時、あらかじめ設定された発信者別のリップアニメーション画像を読み出し、前記抽出されたフォルマント情報の変化に同期してリップシンク・アニメーション画像を生成するリップシンク・アニメーション画像生成手段と、を備えたことを特徴とする携帯電話。

【請求項５】　着信時におけるリップアニメーション画像表示は、ハンズフリー通話モードにおいてのみ有効とするモード制御手段、を備えたことを特徴とする請求項４に記載の携帯電話。

【請求項６】　通常モードにおいてはメールテキストを音声変換して音声出力すると共にリップシンク・アニメーション表示を行い、マナーモードにおいては前記音声出力を禁止してリップシンク・アニメーション表示を行うモード制御手段、を備えたことを特徴とする請求項４に記載の携帯電話。

【請求項７】　アニメーション画像生成装置においてリップシンク・アニメーション画像を生成するアニメーション画像生成方法であって、
音声を分析して母音を特徴づけるフォルマント情報を抽出するステップと、前記抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するステップと、を含むことを特徴とするアニメーション画像生成方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、音声に同期してアニメーション画像のリップ形状が変化する、アニメーション画像生成装置、及び同装置を内蔵した携帯電話、並びにアニメーション画像生成方法に関する。

【０００２】

【従来の技術】パーソナルコンピュータや携帯電話において、音声に同期してアニメーション画像の口形状が変化する技術（以下、リップシンク・アニメーションと称する）を実装したものが存在する。具体的には、特開平６－１６２１６６号公報に開示されている。同公報によれば、音声合成システムまたは実音声から発音情報を抽出し、その発音情報に基づきリップアニメーション画像を制御する技術が開示されている。

【０００３】

【発明が解決しようとする課題】しかしながら上記した技術は、あらかじめプログラムされた内容に基づきリップアニメーション形状が変化するものであり、例えば、携帯電話における着信メールの読み上げ等に同期した実時間でのリップシンク・アニメーションが表示されるものではない。既録音の音声についても同様である。

【０００４】本発明は上記事情に鑑みてなされたものであり、プログラムに依存することなく、実時間音声あるいは既録音の音声からリップシンク・アニメーション画像を生成することができる、アニメーション画像生成装置並びに同装置を内蔵した携帯電話、及びアニメーション画像生成プログラムを提供することを目的とする。

【０００５】

【課題を解決するための手段】上記した課題を解決するために本発明は、音声を分析して母音を特徴づけるフォルマント情報を抽出するフォルマント情報抽出手段と、前記抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するリップシンク・アニメーション画像生成手段と、を備えたことを特徴とする。本発明によれば、リップシンク・アニメーション画像生成手段が、フォルマント情報抽出手段により抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するため、プログラムに依存することなく、既録音、あるいは実時間音声のリップシンク・アニメーション表示を可能としたアニメーション画像生成装置を提供することができる。

【０００６】また、本発明において、前記リップシンク・アニメーション画像生成手段は、前記フォルマント情報抽出手段により抽出される低次のフォルマント周波数の相対的位置関係により母音を識別する母音識別手段と、前記識別された母音に同期してあらかじめ用意されたリップアニメーション画像を選択出力するリップアニメーション画像選択出力手段と、を備えたことを特徴とする。本発明によれば、リップアニメーション画像選択出力手段が、母音識別手段により識別された母音に同期

してあらかじめ用意されたリップアニメーション画像を選択出力するため、音声の定常部位に連動して、既録音、あるいは実時間音声のリップシンク・アニメーション画像表示を可能としたアニメーション画像生成装置を提供することができる。

【０００７】また、本発明において、前記リップシンク・アニメーション画像生成手段は、子音を識別し、前記識別された子音に同期してあらかじめ用意されたリップアニメーション画像を選択出力するリップアニメーション画像選択出力手段、を備えたことを特徴とする。本発明によれば、リップアニメーション画像選択出力手段が、音声の過渡部位である子音を認識して識別された子音に同期してあらかじめ用意されたリップアニメーション画像を選択出力することにより、よりリアルで表現豊かなリップシンク・アニメーション画像表示を可能としたアニメーション画像生成装置を提供することができる。

【０００８】上記した課題を解決するために本発明は、着信音声を分析して母音を特徴づけるフォルマント情報を抽出するフォルマント情報抽出手段と、着信時、あらかじめ設定された発信者別のリップアニメーション画像を読み出し、前記抽出されたフォルマント情報の変化に同期してリップシンク・アニメーション画像を生成するリップシンク・アニメーション画像生成手段と、を備えたことを特徴とする。本発明によれば、リップシンク・アニメーション画像生成手段が、フォルマント情報抽出手段により抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するため、既録音、あるいは実時間音声のリップシンク・アニメーション表示を可能とした携帯電話を提供することができる。

【０００９】また、本発明は、着信時におけるリップアニメーション画像表示は、ハンズフリー通話モードにおいてのみ有効とするモード制御手段、を備えたことを特徴とする。本発明によれば、モード制御手段が、着信時におけるリップアニメーション画像表示をハンズフリー通話モードにおいてのみ有効とするため、通話中に意味のないアニメーション画像表示を無効にすることができ、無駄な電力消費を回避した携帯電話を提供することができる。

【００１０】また、本発明において、通常モードにおいてはメールテキストを音声変換して音声出力すると共にリップシンク・アニメーション表示を行い、マナーモードにおいては前記音声出力を禁止してリップシンク・アニメーション表示を行うモード制御手段、を備えたことを特徴とする。本発明によれば、モード制御手段が、通話モード時、メールテキストを音声変換して音声出力すると共にリップシンク・アニメーション表示を行い、マナーモード時、音声出力を禁止してリップシンク・アニメーション表示を行うため、メールテキストの読み上げをモードによって制御可能な携帯電話を提供することができる。

【００１１】上記した課題を解決するために本発明は、アニメーション画像生成装置においてリップシンク・アニメーション画像を生成するアニメーション画像生成方法であって、音声を分析して母音を特徴づけるフォルマント情報を抽出するステップと、前記抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するステップと、を含むことを特徴とする。本発明によれば、抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するため、既録音、あるいは実時間音声のリップシンク・アニメーション表示が可能になる。

【００１２】

【発明の実施の形態】図１は、本発明のアニメーション画像生成装置の内部構成を機能展開して示したブロック図である。以下に示す各ブロックは、具体的にはＣＰＵならびにメモリを含む周辺ＬＳＩで構成され、ＣＰＵがメモリに記録されたプログラムを読み出し実行することにより、そのブロックが持つ機能を実現するものである。

【００１３】本発明のアニメーション画像生成装置は、フォルマント情報抽出部１と、リップシンク・アニメーション（ＬＳＡ）画像生成部２で構成される。フォルマント情報抽出部１は、音声を分析して母音を特徴づけるフォルマント情報を抽出する機能を持つ。ＬＳＡ画像生成部２は、フォルマント情報抽出部１で抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成する機能を持ち、母音識別部２１と、子音識別部２２と、パラメータ演算部２３と、ＬＡ画像選択出力部２４と、ＬＡ画像データベース（ＤＢ）２５で構成される。

【００１４】母音識別部２１は、フォルマント情報抽出部１により抽出される低次のフォルマント周波数の相対的位置関係により母音を識別する機能を持ち、子音識別部２２は、子音を識別する機能を持つ。また、パラメータ演算部２３は、母音識別部２１、子音識別部２２によって識別された母音、子音パラメータを加工してフォルマント合成を行い、予め多数のＬＡ画像が蓄積されたＬＡ画像ＤＢ２５から適切なＬＡ画像を選択出力するためのデータを生成する機能を持つ。ＬＡ画像選択出力部２４は、パラメータ演算部２３によって出力されるデータに基づき、識別された母音、あるいは子音との組み合わせに同期してあらかじめ用意されたＬＡ画像を選択出力する機能を持つ。

【００１５】図２は、図１に示すアニメーション画像生成装置の動作を説明するために引用したフローチャートであり、具体的には、本発明のアニメーション画像生成プログラムの処理手順を示す。以下、図２に示すフローチャートを参照しながら図１に示すアニメーション画像

生成装置の動作について説明する。

【００１６】普通、母音には数個のフォルマントがあって、周波数の低いほうから第１、第２、第３……フォルマントと呼ばれる。フォルマントは、発声者、性別、年齢等により大幅に変動があり、発声時、前後につなげて発声される音素の影響を受けてその周波数が変動する性質を持つ。母音を特徴付けるフォルマントは、低次のフォルマントであって、特に、第１および第２フォルマントの寄与が大きい。図３、図４に、ＣＳＭ（複合正弦波モデル）分析における分析結果を、それぞれ男性と女性の実測値（縦軸に５母音別の利得、横軸に周波数を示す）で示している。図中、（ａ）は、基本ピッチ成分、（ｂ）は、第１フォルマント、（ｃ）は第２フォルマントを示す。

【００１７】そこで、フォルマント情報抽出部１は、音声を分析して母音を特徴づけるフォルマント情報を抽出し（ステップＳ２１）、抽出されたフォルマント情報をＬＳＡ画像生成部２へ引き渡し、フォルマント分析および分析結果に基づくＬＳＡ画像生成を行う。ここでは、ＣＳＭによりフォルマント分析を行うこととしたが、他に、ＦＦＴ（Fast Fourier Transform）によりスペクトル包絡を得る他に、ＬＰＣ（線形予測）に基づく方法がある（ステップＳ２２）。これらフォルマント分析法については周知の技術であるため、ここでの説明は省略する。

【００１８】リップシンクは、主に５母音（ア・イ・ウ・エ・オ）のタイプによりあらかじめＬＡ画像ＤＢ２５に用意されたリップ画像を切替えることにより実現される。これは、音声の定常部位に連動させる場合であって、母音識別部２１で、第１フォルマントと第２フォルマントの相対位置関係により母音を識別し（ステップＳ２３）、パラメータ演算部２３で合成して得られるデータに基づき、ＬＡ画像選択出力部２４がＬＡ画像ＤＢ２５をアクセスすることによって得られる（ステップＳ２４、Ｓ２５）。

【００１９】なお、音声の過渡部、すなわち、子音を認識して上記したリップシンクを制御すれば、よりリアルで表現性豊かなリップアニメーション画像の生成が可能である。このときの子音は、子音識別部２２で検出される。また、音声を分析して母音、子音を検出する技術、ならびに、検出したパラメータを加工してフォルマント合成する技術は、特許２７５４９６５号に詳細に開示されているため、ここでの説明は省略する。

【００２０】上記した本発明のアニメーション画像生成装置の携帯電話への応用が図５に示されている。図５において、１０は携帯電話本体であり、図示せぬアンテナ、変復調部、ＣＯＤＥＣ他、マイクロプロセッサ内蔵の制御部で構成される。１１は上記したアニメーション画像生成装置である。１２はテキスト—音声変換装置であり、メールテキストを読み上げるときに使用される。上記した携帯電話本体１０、アニメーション画像生成装置１１、テキスト—音声変換装置１２は、いずれもモード制御部１３によってその動作モードが制御される。

【００２１】ここで、動作モードとして用意されるものに、通常モードの他に、ハンズフリー（ＨＦ）モード、マナーモードがある。すなわち、通話時における発信者のリップアニメーション表示は、モード制御部１３のコントロールにより、ハンズフリー通話時にのみ有効とする。これは、受話口に耳をあてて通話しているときにアニメーション表示しても意味がないことへの対応であり、この場合無駄な表示動作が省けるため、携帯電話のバッテリ延命化に寄与する。なお、発信者別のアニメーション画像をあらかじめＤＢ化しておき、着信時、発信者のアニメーション画像が出現してそのリップ形状を制御することによりリアリティが増す。

【００２２】一方、メールテキストの読み上げ時、通常モードにおいては、テキスト—音声変換装置１２により入力テキストが音声変換され、更に読み上げ音声がリップシンク・アニメーション画像に変換され表示される。この場合、テキスト表示、読み上げ音声出力、リップアニメーション表示が共になされる。これに対し、マナーモードが設定されていた場合、テキスト—音声変換装置１２により入力テキストが音声変換されるが、モード制御部１３により音声出力は禁止される。また、読み上げ音声は、アニメーション画像生成装置１１によりリップシンク・アニメーション画像に変換され表示される。この場合、テキスト表示、リップアニメーション表示は共になされる。なお、テキスト表示は、音声またはリップアニメーション画像に同期して色替えまたはスクロール表示されることとする。

【００２３】以上説明のように、本発明によれば、フォルマント情報に基づく母音情報を得てリップ形状を制御することにより、プログラムに依存することなく、実時間音声あるいは既録音の音声からリップシンク・アニメーション画像を生成することが可能となる。なお、上述の実施の形態では、アニメーション画像生成装置の応用例として携帯電話についてのみ示したが、ＰＤＡ（Personal Digital Assistants）等、携帯電話機能付き情報機器にも同様に応用可能である。また、図１に示すフォルマント情報抽出部１、ＬＳＡ画像生成装置２、母音識別部２１、子音識別部２２、パラメータ演算部２３、ＬＡ画像選択出力部のそれぞれで実行される手順をコンピュータ読取り可能な記録媒体に記録し、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより、本発明のアニメーション画像生成装置が実現されるものとする。ここでいうコンピュータシステムとは、ＯＳや周辺機器等のハードウェアを含む。

【００２４】更に、「コンピュータシステム」は、ＷＷＷシステムを利用している場合であれば、ホームページ

提供環境（あるいは表示環境）も含むものとする。更にまた、「コンピュータ読取り可能な記録媒体」とは、ＲＯＭの他に、フレキシブルディスク、光磁気ディスク、ＣＤ—ＲＯＭ等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のシステムやクライアントとなるコンピュータシステム内部の揮発性メモリ（ＲＡＭ）のように、一定時間プログラムを保持しているものも含むものとする。

【００２５】更にまた、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。更にまた、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【００２６】以上、この発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

【００２７】

【発明の効果】以上説明のように本発明によれば、抽出されたフォルマント情報の変化に追随してリップシンク・アニメーション画像を生成するため、プログラムに依存することなく、既録音、あるいは実時間音声のリップシンク・アニメーション表示が可能になる。また、母音の他に、音声の過渡部位である子音を認識し、識別された子音に同期してあらかじめ用意されたリップアニメーション画像を選択出力することにより、よりリアルで表現豊かなリップシンク・アニメーション画像表示を可能とすることができる。また、本発明を携帯電話に応用することにより、メールテキストの読み上げ等、動作モードに従うリップシンク・アニメーション表示を実現できる。

【図面の簡単な説明】

【図１】　本発明のアニメーション画像生成装置の内部構成を機能展開して示したブロック図である。

【図２】　図１に示すアニメーション画像生成装置の動作を説明するために引用したフローチャートである。

【図３】　本発明実施形態の動作を説明するために引用したグラフであり、ＣＳＭ分析結果である実測値を示す。

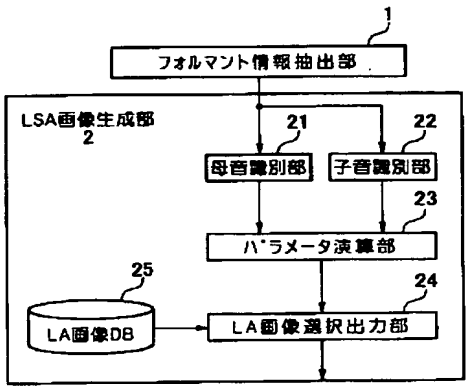
【図４】　本発明実施形態の動作を説明するために引用したグラフであり、ＣＳＭ分析結果である実測値を示す。

【図５】　本発明の形態電話への応用例を示すブロック図である。

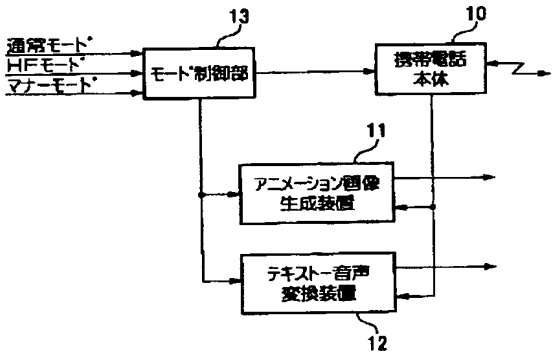
【符号の説明】

１…フォルマント情報抽出部、２…リップシンク・アニメーション（ＬＳＡ）画像生成部、１０…携帯電話本体、１１…アニメーション画像生成装置、１２…テキスト—音声変換装置、１３…モード制御部、２１…母音識別部、２２…子音識別部、２３…パラメータ演算部、２４…リップアニメーション（ＬＡ）画像選択出力部、２５…リップアニメーション（ＬＡ）画像ＤＢ。

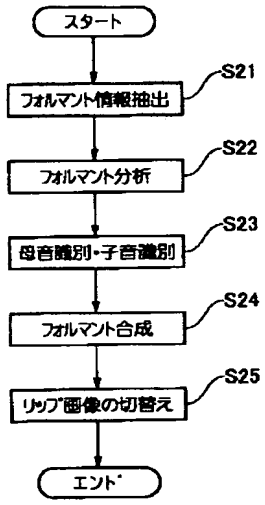
【図１】



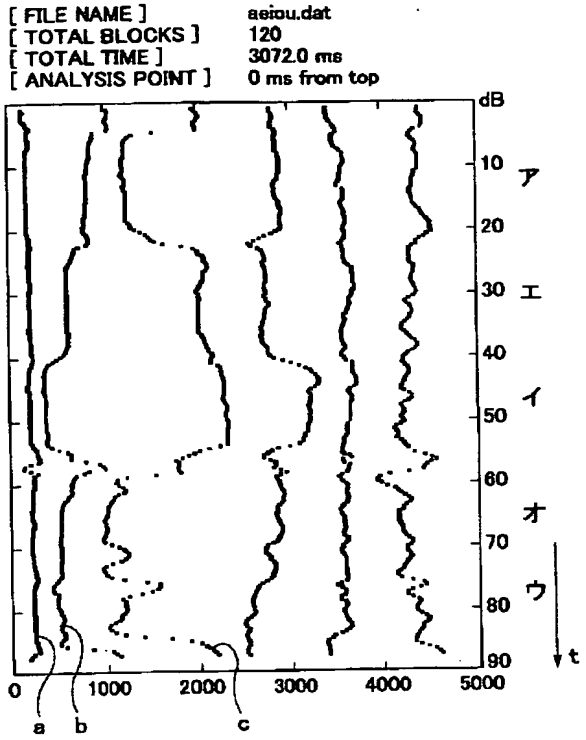
【図５】



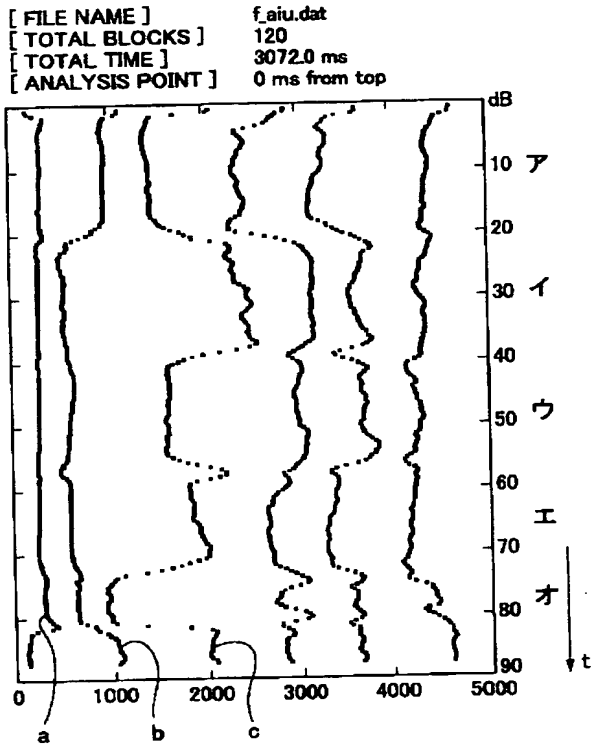
【図２】


スタート
フォルマント情報抽出
S21
フォルマント分析
S22
母音識別・子音識別
S23
フォルマント合成
S24
リップ画像の切替え
S25
エント'

【図３】


[FILE NAME] aeiou.dat
[TOTAL BLOCKS] 120
[TOTAL TIME] 3072.0 ms
[ANALYSIS POINT] 0 ms from top
dB
10
20
30
40
50
60
70
80
90
ア
エ
イ
オ
ウ
t
0
1000
2000
3000
4000
5000
a
b
c

【図４】



フロントページの続き

(51) Int. Cl.[7]	識別記号	F I	テーマコード（参考）
		Ｇ１０Ｌ　3/00	５５１Ａ